

# **Independent Study Module(ISM) Report: CS-IS-4074-1**

**Submitted by:**

Santosh Adhikari

This report outlines the tasks and responsibilities undertaken as part of my ISM (CS-IS-4074-1): Evaluation of vision transformer models for gigapixel histopathology images. [GitHub Link](#)

**Departments:**

Department of Computer Science  
Koita Center For Digital Health Ashoka (KCDH-A)

**Supervised by:**

Professor Rintu Kutum

May 17, 2025

# 1 Breast Cancer

According to the WHO Report of 2022, there were 2.3 million new cases annually with 670,000 deaths in 2022 due to breast cancer. As of 2022, India ranks highest in the number of estimated breast cancer deaths (98,337) while 1 in 56 people and 1 in 29 women received a diagnosis as of 2020 data.

## 1.1 Triple Negative Breast Cancer (TNBC)

Triple Negative Breast Cancer (TNBC) is an aggressive type of cancer where its cells lack the Human Epidermal Growth Factor-2 (HER2) antigen and Hormone Receptors Estrogen (ER) and Progesterone (PR). TNBC is referred to as triple negative, meaning that its cells are ER-ve, PR-ve, and HER2-ve. The existence of these molecular markers is determined using various immunohistochemistry techniques.

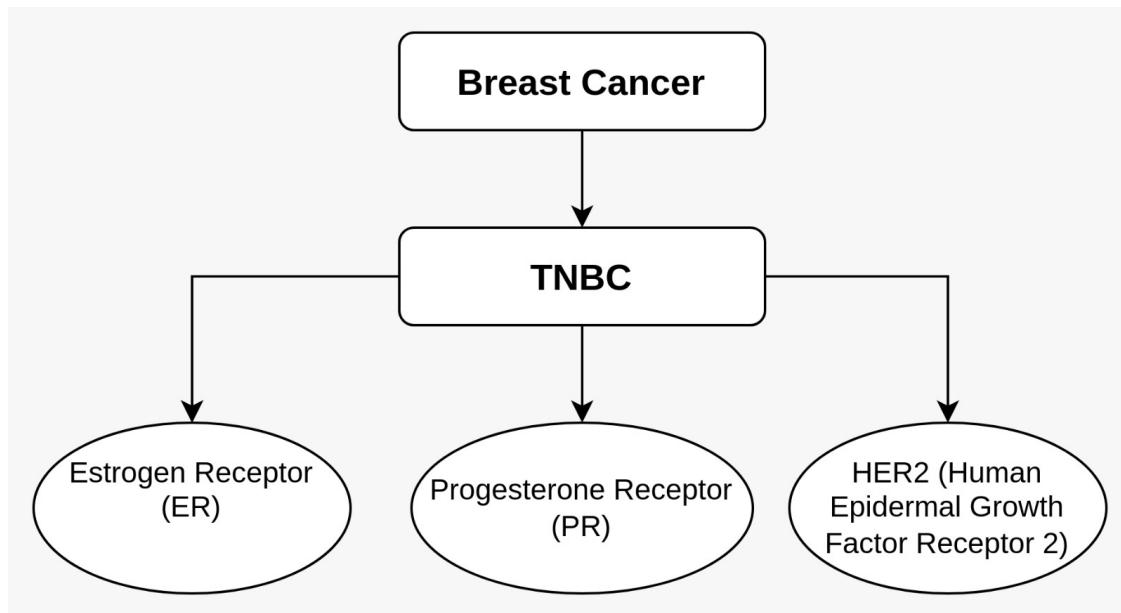


Figure 1: Breast Cancer, TNBC, Estrogen Receptor (ER), Progesterone Receptor (PR), HER2 (Human Epidermal Growth Factor Receptor 2)

# 2 Triple Negative Breast Cancer (TNBC) Dataset : IISER, Pune

- Dataset storage system's information: Wellcome Trust 1
- IP Address : 10.1.20.243
- Storage Path: /media/network/04-tnbc/tnbc

## 2.1 Dataset Summery

The dataset comprises Whole Slide Images (WSIs) of breast cancer tissue samples collected across 10 experimental rounds, each labeled with specific molecular subtypes: ER (Estrogen Receptor positive), HER2 (Human Epidermal Growth Factor Receptor 2) and TNBC (Triple Negative Breast

Cancer). The distribution of samples varies across rounds, with TNBC being most prominent in Rounds 1 and 6, HER2 in Rounds 1 and 2, and ER mainly in Rounds 1 to 3 and 8. HER3 is minimally represented, appearing in only one sample. Each WSI is scanned at 40X magnification, allowing for high-resolution pathological analysis.

Each patient sample is annotated with both its molecular subtype and the specific histological or immunohistochemical stains applied. Common stains include HnE (Hematoxylin and Eosin), used for general tissue structure visualization, as well as marker specific stains like Ki-67 (proliferation), CD31 (angiogenesis), AR (Androgen Receptor), CK5-6, Vimentin, and YAP. Subtype-stain combinations vary: ER samples often have HnE, Ki-67, and YAP; TNBC samples are typically stained with a broader panel including AR, and Vimentin; HER2 cases frequently include HER3 and Ki-67 among others. This multimodal dataset is well-suited for tasks such as subtype classification, biomarker discovery, and development of machine learning models that leverage stain-specific features.

## 2.2 Data Storage Information

### a) OPTRASCAN\_IISER\_Round\_1

The Dataset contains the TNBC, ER and HER2 whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 80
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_1/IISER\_Box\_1/TIFF/
- Box 2 :
  - WSIs count: 53
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_1/IISER\_Box\_2/TIFF/
- Box 3 :
  - WSIs count: 49
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_1/IISER\_Box\_3/TIFF/
- Box 4 :
  - WSIs count: 92
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_1/IISER\_Box\_4/TIFF/
- Box 5 :
  - WSIs count: 66
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_1/IISER\_Box\_5/TIFF/
- Box 6 :
  - WSIs count: 85
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_1/IISER\_Box\_6/TIFF/
- Box 7 :

- WSIs count: 44
- Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_1/IISER\_Box\_7/TIFF/

### **b) OPTRASCAN\_IISER\_Round\_2**

The Dataset contains the TNBC, ER and HER2 whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 73
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_2/IISER\_Box\_1/Tif/
- Box 2 :
  - WSIs count: 53
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_2/IISER\_Box\_2/Tif/
- Box 3 :
  - WSIs count: 65
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_2/IISER\_Box\_3/Tif/
- Box 4
  - WSIs count: 91
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_2/IISER\_Box\_4/Tif/
- Box 5 :
  - WSIs count: 60
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_2/IISER\_Box\_5/Tif/
- Box 6 :
  - WSIs count: 60
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_2/IISER\_Box\_6/Tif/
- Box 7 :
  - WSIs count: 54
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_2/IISER\_Box\_7/Tif/

### **c) OPTRASCAN\_IISER\_Round\_3**

The Dataset contains the TNBC, ER and HER2 whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 54
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_3/IISER\_BOX1\_Tiff
- Box 2 :

- WSIs count: 56
- Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_3/IISER\_BOX2\_Tiff

#### d) OPTRASCAN\_IISER\_Round\_6

The Dataset contains the TNBC whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 60
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_6/BOX1/tif
- Box 2 :
  - WSIs count: 60
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_6/BOX2/tif
- Box 3 :
  - WSIs count: 60
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_6/BOX3/tif

#### e) OPTRASCAN\_IISER\_Round\_7

The Dataset contains the TNBC whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 36
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_7/BOX1/tif
- Box 2 :
  - WSIs count: 36
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_7/BOX2/tif
- Box 3 :
  - WSIs count: 36
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_7/BOX3/tif

#### f) OPTRASCAN\_IISER\_Round\_8

The Dataset contains the ER and HER2 whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 59
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_8/TIFF

#### f) OPTRASCAN\_IISER\_Round\_9

The Dataset contains the TNBC and HER2 whole slide images with multiple stains. This dataset contains two WSIs whose subtype is missing.

- Box 1 :
  - WSIs count: 40
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_9/BOX1/TIF
- Box 2 :
  - WSIs count: 27
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_9/BOX2/TIF

#### g) OPTRASCAN\_IISER\_Round\_10

The Dataset contains the TNBC whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 16
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_10/TIF

#### h) OPTRASCAN\_IISER\_Round\_11

The Dataset contains the HER2 whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 37
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_11/TIF

#### f) OPTRASCAN\_IISER\_Round\_12

The Dataset contains the HER2 whole slide images with multiple stains.

- Box 1 :
  - WSIs count: 37
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_12/IISER\_BOX1/Tiff
- Box 2 :
  - WSIs count: 37
  - Path: /media/network/04-tnbc/tnbc/OPTRASCAN\_IISER\_Round\_12/IISER\_BOX1/Tiff

### 2.3 Analysis on the Dataset

This section presents a visual analysis of the IISER TNBC dataset, highlighting distributions and counts across various features such as rounds, stains, subtypes, and patient IDs.

### 2.3.1 Total File Counts and Patient's Information

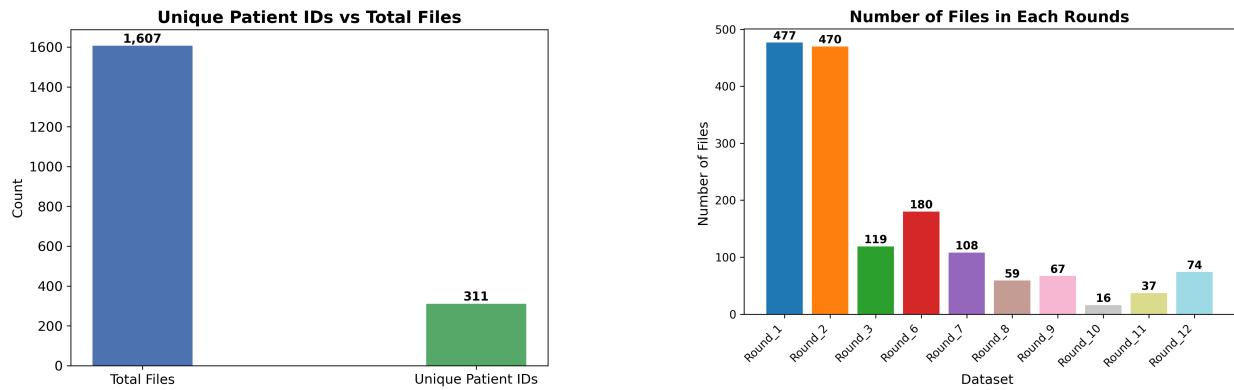


Figure 2: Left: Total Unique Patient IDs, Right: Total Number of Files in Each Round

This left bar plot compares the total number of files in the dataset against the number of unique patient IDs.

- There are **1,607 Total Files**.
- These files correspond to **311 Unique Patient IDs**.

This indicates that there are multiple files per patient while some patients have a single file. This could be due to multiple biopsies, different stains applied to samples from the same patient. For instance, **Patient ID: 2** has just as single file with Esterogene(ER) subtype with H&E staining while the **Patient ID: 3** has TNBC subtype but with 8 different staining (AR, CD31, CK5-6, HER3, HnE, Ki-67, Vimentin, YAP).

This right bar chart shows the total number of WSI files processed or available in each experimental round.

- **Round\_1** (477 files) and **Round\_2** (470 files) have the highest number of files.
- **Round\_6** (180 files), **Round\_3** (119 files), and **Round\_7** (108 files) also contribute a significant number of files.
- The remaining rounds (**Round\_8**, **Round\_9**, **Round\_10**, **Round\_11**, **Round\_12**) have fewer files, ranging from 16 to 74.

### 2.3.2 File Count Based on Sub type Information

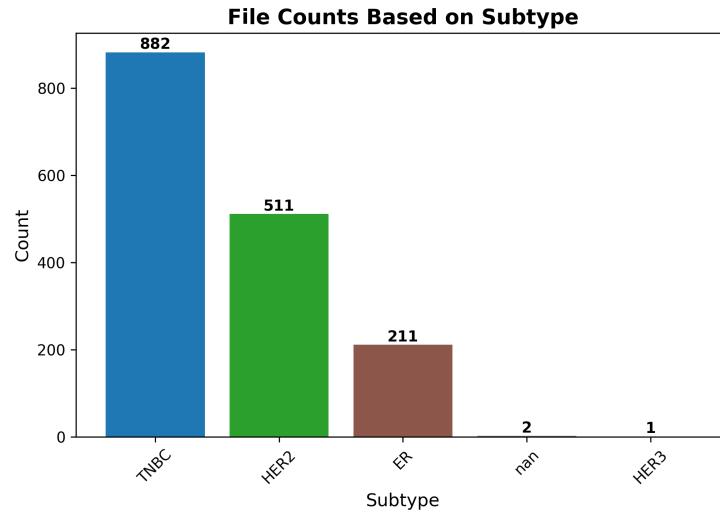


Figure 3: Total Files based on Sub type Information

The plot shows an aggregated view of file counts for each cancer subtype across the entire dataset.

- **TNBC** is the most represented subtype with 882 files.
- **HER2** positive samples are the second most common, with 511 files.
- **ER** positive samples have 211 files.
- A small number of files are categorized as “nan” (2 files) or **HER3** (1 file) at the subtype level.

### 2.3.3 File Distribution based on Sub Type in Each Round

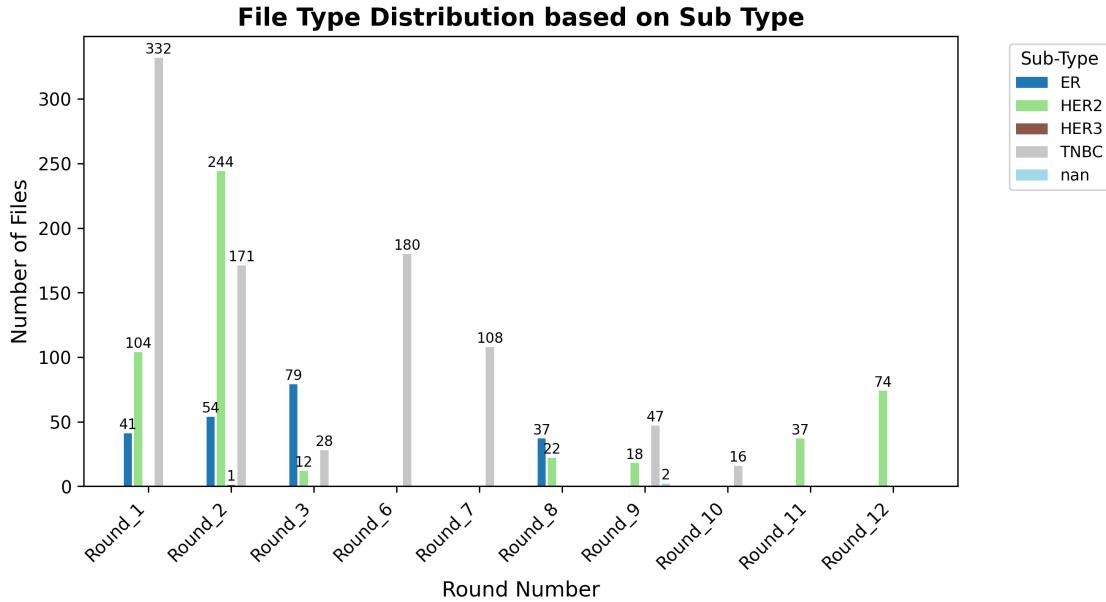


Figure 4: File Distribution based on Sub Type

The plot shows the distribution of file subtypes (ER, HER2, HER3, TNBC, nan) across different experimental rounds (Round\_1 through Round\_12).

- **Round\_1** shows a mix, with a notable number of HER2 (104) and TNBC (41) files.
- **Round\_2** is predominantly composed of TNBC files (332), with a significant count of HER2 files (244) and a smaller number of ER files (54).
- **Round\_3** primarily contains HER2 files (79) and a smaller amount of TNBC (28) and ER (12) files.
- **Round\_6** and **Round\_7** consist exclusively of TNBC files, with counts of 180 and 108 respectively.
- **Round\_8** shows a mix of ER (37), HER2 (22), and TNBC (37) files.
- **Round\_9** is dominated by TNBC files (47), with a few HER2 (18) and ER (2) files.
- **Round\_10** and **Round\_11** mainly contain TNBC (16) and HER2 (37) files respectively.
- **Round\_12** consists solely of HER2 files (74).

This distribution indicates different breast cancer subtypes across the experimental rounds, with TNBC and HER2 being prominent. The “nan” category represents files where subtype information was not available.

### 2.3.4 File Count Based on Stain Information

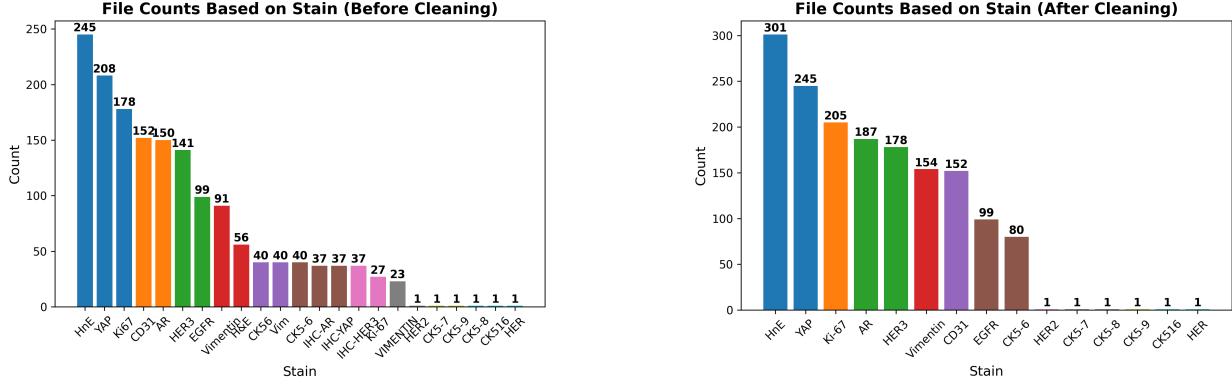


Figure 5: Left: Before cleaning based on stain, Right: After cleaning based on stain

The first plot shows the raw counts of files for each type of stain applied before any data cleaning.

- The most frequent stains are **HnE** (245 files), **YAP** (208 files), and **Ki67** (178 files).
- Other common stains include **CD31**, **AR**, **HER3**, **EGFR**, and **Vimentin**, with counts ranging from 152 down to 91.
- Several **CK5/6** combinations and individual IHC stains (like **IHC-YAP**, **IHC-HER3**) also appear, though in smaller quantities (23 to 40 files).
- A long tail of stains with very few files (count of 1) is visible, such as **VIMENTIN HER2**, **CK5-7**, **CK5-9**, **CK5-8**, **CK516**, and **HER**, suggesting either highly specific experiments or potential inconsistencies in stain labeling that might require cleaning.

Following data cleaning, this bar chart shows the revised counts of files per stain.

- **HnE** remains the most prevalent stain (301 files), followed by **YAP** (245 files) and **Ki-67** (205 files).
- **AR** (187 files) and **HER3** (178 files) also show significant counts.
- Stains like **Vimentin**, **CD31**, **EGFR**, and **CK5-6** have counts between 80 and 154.
- The number of unique stains with very low counts (e.g., count of 1 for **HER2**, **CK5-7**, **CK5-8**, **CK5-9**, **CK516**, **HER**) has been reduced or consolidated compared to the “before cleaning” state, indicating successful normalization or removal of erroneous/rare entries.

### 2.3.5 Top 10 Subtype and Stain Combinations Across Patients

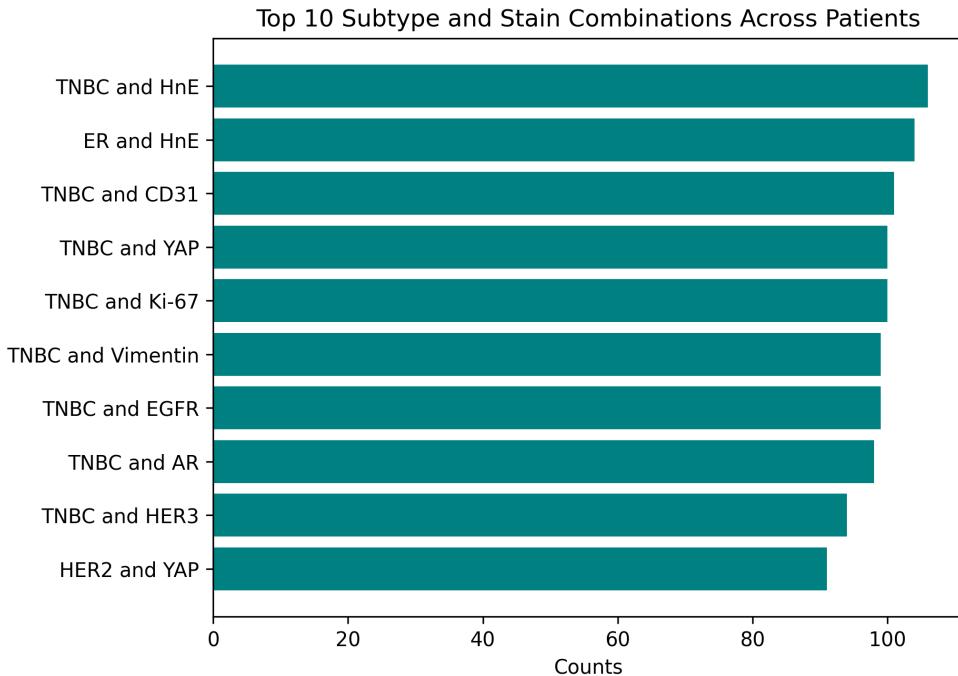


Figure 6: Top 10 Subtype and Stain Combinations

This horizontal bar chart displays the ten most frequent combinations of cancer subtype and applied stain found across unique patients.

- The most common combination is **TNBC and HnE**, appearing in over 100 unique patients.
- **ER and HnE** is the second most frequent.
- Several other TNBC combinations with IHC stains like **CD31**, **YAP**, **Ki-67**, **Vimentin**, **EGFR**, **AR**, and **HER3** are also prominent, each occurring in 90–100 unique patients.

## 3 Analysis of Segmentation Techniques

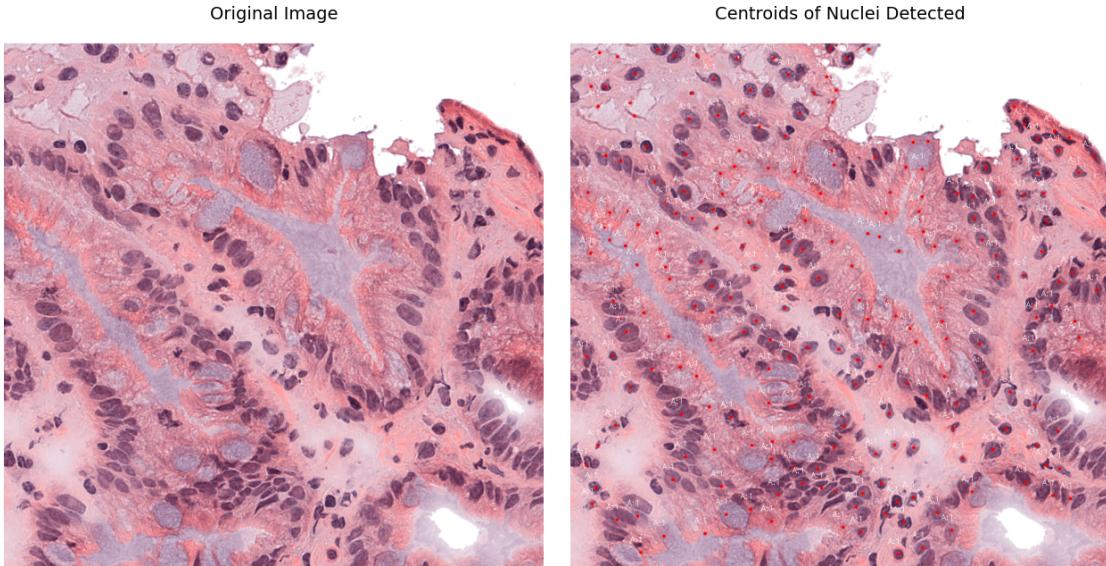
Building upon the foundational work with datasets and software tools, a significant task involved applying and evaluating machine learning models for tasks such as cell type prediction and segmentation.

### 3.1 Watershed Algorithm

The watershed algorithm is a popular method for computational image segmentation that has its roots in mathematical morphology. The algorithm views a grayscale image as a topographic surface, with elevations represented by pixel intensities. The algorithm was first presented by Vincent and Soille and views a grayscale image as a topographic surface, with elevations represented by pixel intensities. The main concept is to treat local intensity minima as water sources that 'flood' the topography of the image. The image is essentially divided into discrete regions as the water fills the

landscape and creates boundaries where floods from various minima converge. These lines define the areas or items in the picture and are also referred to as watershed lines. In biomedical image analysis, the watershed algorithm has been a popular method because of its easy-to-understand structure and computational efficiency, especially for detecting nuclei in histopathological images. Frequently used to improve feature differentiation in an image, the algorithm comes after preprocessing steps like gradient calculation or marker based labeling. For instance, marker controlled watershed segmentation reduces oversegmentation by using predefined markers to direct the flooding process, increasing segmentation accuracy.

The algorithm was implemented from the scratch to test how effectively it is able to segment the tissue patch and confirm the presence of nucleus. The image is first preprocessed using Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve local contrast. Gaussian filtering is then applied to minimize noise while preserving edge information. After preprocessing, the image is binarized using Otsu's thresholding technique, which makes the background and nuclei distinct. A topographic map with peaks at the centers of nuclei is produced by the distance transform, which is used in the watershed segmentation process to find markers. Each nucleus forms a catchment basin in the watershed algorithm, which uses these peaks as markers and views the image as a topographic surface.



(a) Left: Original WSI Patch, Right: Nuclei detection using the Watershed Algorithm

The watershed approach's sensitivity to image noise and staining variations is one of its main drawbacks. Variations in sample preparation or imaging conditions frequently result in inconsistent staining intensities particularly those stained with H&E. This discrepancy leads to incorrect segmentations and watershed boundaries. The algorithm also has trouble in cellular environments that are densely packed as overlapping nuclei and crowded tissue regions cause oversegmentation. This causes the splits in single objects into multiple fragments.

### 3.2 Mask R-CNN Model

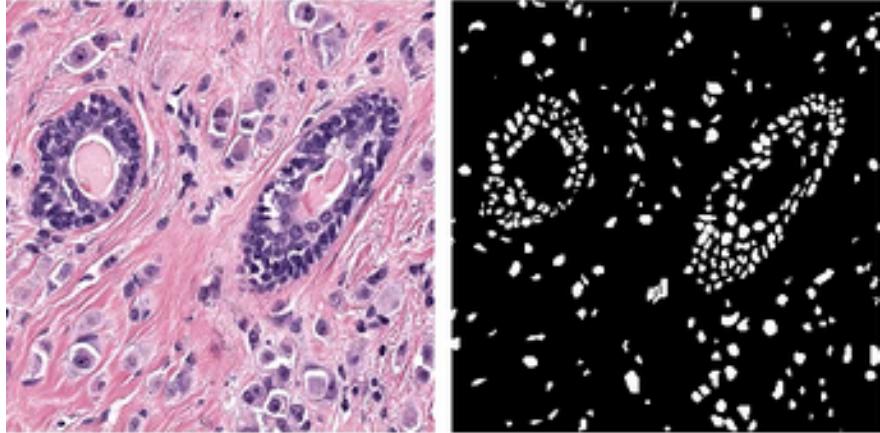
Mask R-CNN is a model that can perform both instance segmentation and object detection at the same time. Its architecture can handle complex visual tasks like segmenting cellular structures that

are morphologically variable and densely packed.

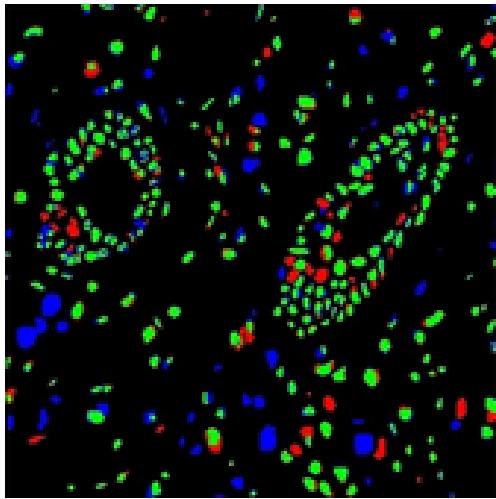
The architecture of Mask R-CNN mainly contains three main stages:

- Backbone Network for Feature Extraction
- Region Proposal Network (RPN)
- Region of Interest (RoI) Align Layer

A deep convolutional neural network like ResNet-50 or ResNet-101 serves as the backbone network. To effectively extract multi-scale features from the input image, a Feature Pyramid Network (FPN) is added. The creation of feature maps that extract contextual and spatial information from the input image is the responsibility of this step. Where objects (like nuclei) are likely to be found, the RPN suggests potential regions of interest (RoIs). Predicting bounding boxes and objectness scores for every region entails dragging a tiny network across the feature maps. For cellular image analysis, the RPN is very effective because it functions as a filter, reducing the regions to only those that are most likely to contain nuclei. Mask R-CNN introduces RoI Align, which uses bilinear interpolation to ensure precise spatial mapping. For biomedical images, where small or overlapping nuclei must be distinguished with sub-pixel-level accuracy, this improvement is essential.



Left: H&E stained WSI, Middle: Ground Truth Mask



Mask-R-CNN Prediction

Figure 8: Mask-R-CNN Input and Prediction

### 3.3 Cell-ViT

Cell-ViT is a deep learning architecture based on Vision Transformer that uses enables automated instance segmentation technique for cell nuclei in digital tissue samples. Cell-ViT architecture is based on the UNETR model but has its framework modified for 2D image analysis instead of 3D volumetric image processing. The architecture includes three different multitask output branches. The first branch captures the boundaries and morphological characteristics of each nucleus by predicting a binary segmentation map of all nuclei (Nuclei Prediction or NP). In order to accurately locate and delineate nuclei, the second branch generates horizontal and vertical distance maps (horizontal-vertical prediction, or HV). The third branch predicts nuclei type map (NT) for classification of various nuclei types. This multi-branch architecture improves the model's capacity to manage classification, localization, and segmentation tasks within a single framework.

#### CellViT Architecture:

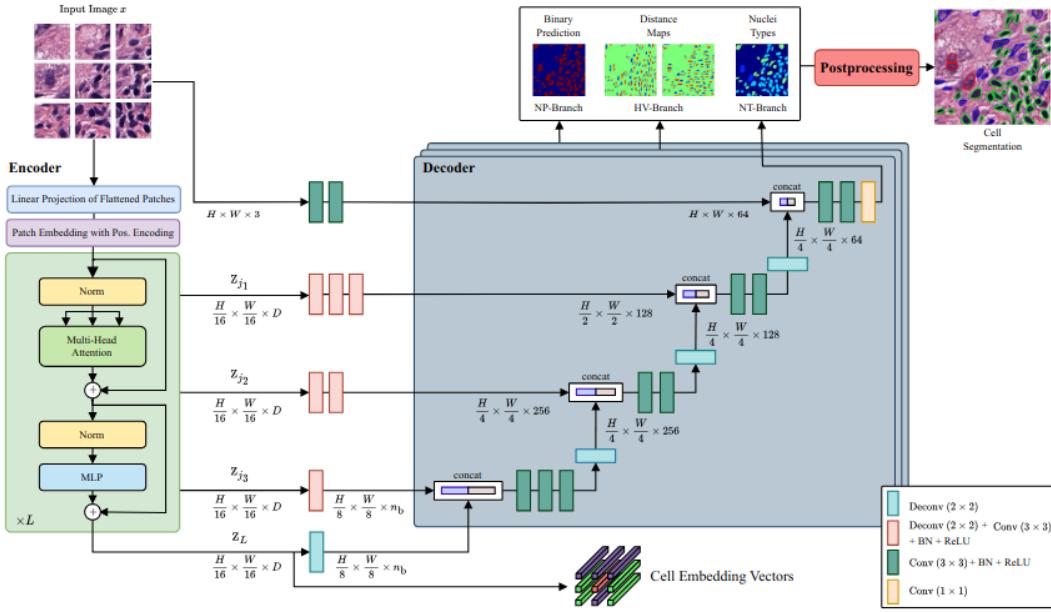


Figure 9: CellViT architecture consisting of a vision transformer encoder connected to multiple decoders via skip connections.

## Our Proposed Methodology:

The methodology for generating a high quality annotated dataset from Whole Slide Images (WSIs) is designed as a systematic, multi-stage pipeline that ensures both efficiency and accuracy in cell-level annotation. The process begins with data harmonization where raw WSIs are processed and converted into a standardized pyramidal TIFF format. This conversion facilitates efficient navigation and access to image data at multiple resolutions which is essential for handling the typically large sizes of WSIs in digital pathology.

Following this, a WSI mask generation step is employed to automatically identify and exclude background regions from the slide. This step ensures that only the biologically relevant tissue regions are retained for analysis, optimizing computational resources and improving model performance by focusing on meaningful areas. The tissue regions identified by the mask are then used to extract image patches of size  $1024 \times 1024$  pixels. To enhance patch quality and consistency, preprocessing steps such as 64-pixel patch overlap, stain normalization, and downsampling are applied. Patch overlap increases the continuity of features across boundaries, while stain normalization addresses variability in staining protocols, making the data more uniform for deep learning. Downsampling reduces the computational burden without significantly compromising image detail. These refined image patches serve as input to the Cell-ViT model, a vision transformer-based architecture specifically tailored for cell-level analysis. The model performs two primary tasks:

- First, it generates cell embeddings, capturing high-dimensional representations of cellular morphology and spatial context and,
- Predicts cell types by generating cell-level annotations.

These predictions are informed by the learned representations and can distinguish among various cell types present in the tissue.

To ensure annotation quality, the generated cell embeddings are made available for expert re-annotation through a dedicated application providing flexibility for domain experts to refine or correct the model's predictions. Additionally, a manual review process is incorporated to validate the automated annotations, allowing human experts to inspect, confirm, or adjust the cell labels as needed. The final stage of the pipeline integrates both the automated predictions and the expert-reviewed annotations to produce a high-quality, robustly annotated dataset. This dataset is suitable for downstream applications such as biomarker discovery, computational pathology research, and training supervised learning models for various diagnostic and prognostic tasks.

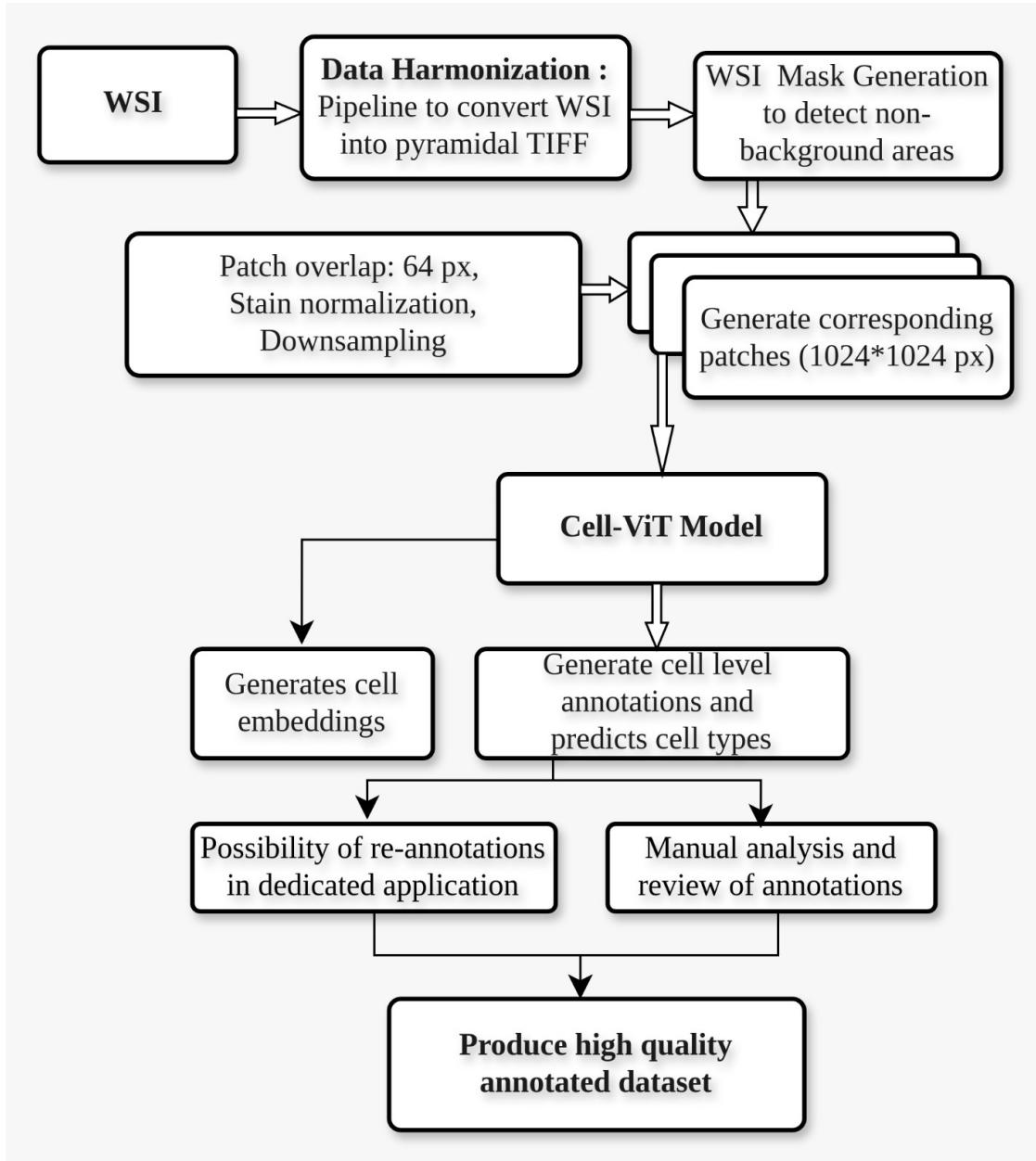


Figure 10: From WSI To High Quality Annotated Dataset

## Data Harmonization

When the data arrived to us, it was in an unprocessed state. Before I could do any calculations, I needed to convert them into a structured format and make the dataset into the input format as expected by the model. It was essential to first understand the metadata and the naming convention used in the laboratory. I created a pipeline that processed every WSI file and produced the necessary JSON-formatted metadata. In addition, I made a structured data structure with slides organized and categorized by patient ID. During the pre-processing stage, the model expects the WSI to be in the pyramidal TIFF format because it uses OpenSlide, a C library that enables it to read WSI. In the pyramidal TIFF format, the file needs to have multiple downscaled resolutions (pyramid levels).

However, since our WSI directly came from the lab, many of them were not in the lab while many of our WSI were not in compatible format. In order to solve this specific problem, I wrote a script using VIPS, a high-performance image processing library, to transform the problematic TIFF into a pyramidal TIFF format. Accordingly, I created a pipeline to pre-process our dataset and make it compatible for the input as required by the Cell-ViT model.

## Main Technical Specifications used for Inference

- Base Magnification: 40.0x
- Patch Size: 1024 pixels
- Patch Overlap: 64 pixels
- Stain Normalization: Applied

For a given patch at position  $(i, j)$  with the detected cell at coordinate  $(x_{\text{patch}}, y_{\text{patch}})$ :

a) **Patch Coordinate:**

The top-left corner of the patch corresponds to the WSI coordinates:

$$(i \times 1024, j \times 1024)$$

b) **Convert Patch Coordinates to WSI Coordinates:**

WSI coordinates  $(x_{\text{wsi}}, y_{\text{wsi}})$  can be calculated as:

$$\begin{aligned}x_{\text{wsi}} &= (i \times 1024 + x_{\text{patch}}) \\y_{\text{wsi}} &= (j \times 1024 + y_{\text{patch}})\end{aligned}$$

If we consider overlapping of patches with overlap percentage of  $z\%$ :

$$\text{stride} = 1024 \times (1 - z\%)$$

E.g., if overlap = 6.25%,

$$\text{stride} = 1024 \times (1 - 0.0625) = 1024 \times 0.9375 = 960$$

Thus, the patch coordinate with respect to WSI coordinate:

$$\begin{aligned}x_{\text{wsi}} &= (i \times \text{stride} + x_{\text{patch}}) \\y_{\text{wsi}} &= (j \times \text{stride} + y_{\text{patch}})\end{aligned}$$

Alternatively, for the specific example with a stride of 960:

$$x_{\text{wsi}} = (i \times 960 + x_{\text{patch}})$$

$$y_{\text{wsi}} = (j \times 960 + y_{\text{patch}})$$

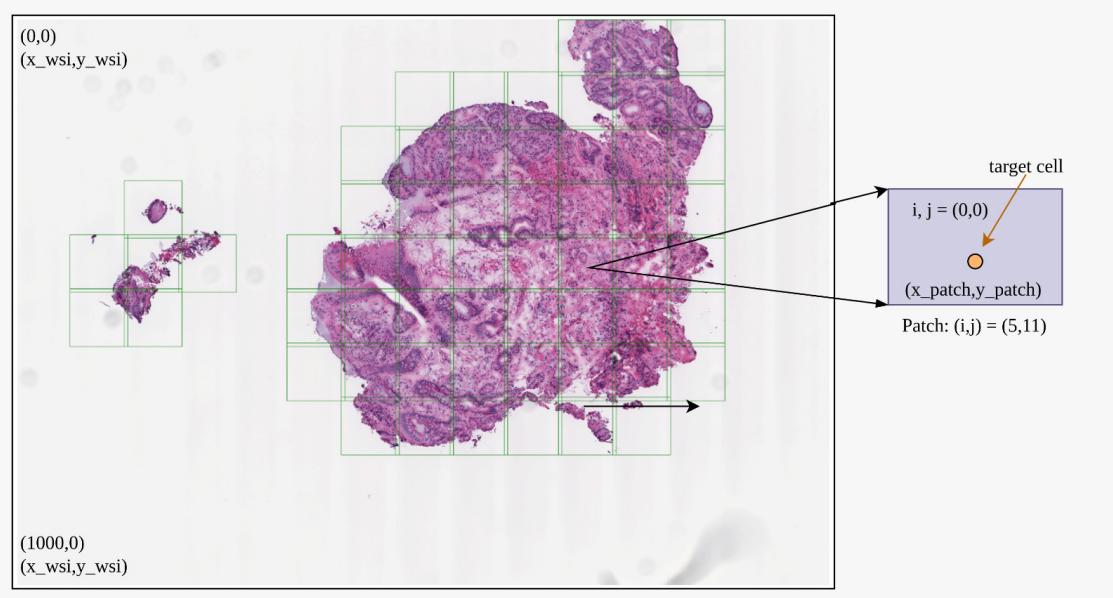


Figure 11: CellViT: Mapping of Local coordinates wrt Global Coordinates of WSI

### Inference: CellViT Model and CellViT-Plus Model

The inference was carried out on the all TNBC H&E stained whole slide images as part of above mentioned IISER Dataset. The Inference was done in the following system :

System Information : RF Grant-2.

#### a) OPTRASCAN\_IISER\_Round\_1

- WSIs count: 80
- Raw Data Path: /media/network/hdd/santosh/OPTRASCAN\_IISER\_Round\_1/IISER\_BOX\_1
- Inference Results : /media/network/hdd/santosh/OPTRASCAN\_IISER\_Round\_1

#### b) OPTRASCAN\_IISER\_Round\_2

- WSIs count: 65
- Raw Path: /media/network/hdd/santosh/OPTRASCAN\_IISER\_Round\_2/IISER\_BOX\_3
- Inference Results : /media/network/hdd/santosh/OPTRASCAN\_IISER\_Round\_2

**Model Weights:** The pretrained model weights were used for inference. The model checkpoint I used for inference can be downloaded from [here](#).

**Input:** The input is a Whole Slide Image (WSI) of a tissue sample. The WSI first undergoes a preprocessing pipeline consisting of the following steps:

- WSI Mask Generation: A WSI mask is generated to identify all non-background patches. At this stage, no patches are extracted.
- Division into Overlapping Grids: The WSI mask is divided into overlapping grids, which are used to generate patches. This process extracts smaller patches from the WSI that serve as inputs for the model.
- Patch Extraction: The WSI is divided into smaller overlapping patches, each of dimension  $1024 \times 1024$ , with a constant overlap of 64 pixels on all sides of the square.

## Result : Cell Type Classification

The model is able to identify and then classify cells into six different categories:

1. **Background: Non-cellular regions.** This region appears as empty or unstained regions with no visible cellular structures and typically lacks texture or color variation.
2. **Neoplastic: Potentially Malignant or Benign.** They are irregularly shaped cells with high nuclear-to-cytoplasmic ratio. They contain dense and darkly stained nuclei, often showing variability in size and shape (pleomorphism). They usually form disorganized patterns in the region where they are present.
3. **Epithelial: Surface and glandular tissue cells.** They are uniform, polygonal cells arranged in cohesive sheets with moderately sized nuclei with clear boundaries.
4. **Inflammatory: Immune cells and inflammatory tissue components.** They are characterized by small, round cells with densely stained nuclei and little visible cytoplasm. They are found scattered or clustered near regions of tissue injury or inflammation.
5. **Connective: Cells associated with connective tissue.** They are spindle-shaped cells with elongated nuclei and are found within extracellular matrix regions. They usually have lighter staining, with sparse and elongated cell arrangements.
6. **Dead: Deceased or degenerating cells.** They are cells with fragmented or faded nuclei (pyknosis, karyorrhexis) and cytoplasm appears to be fragmented.

## Annotation Data Structure

### JSON File Analysis

The JSON annotation file provides information about the cellular detection, including:

- WSI Metadata: Captures imaging parameters and tile configuration.
- Processed Patches: Systematic grid of analyzed image patches.
- Type Mapping: Numerical encoding of cell types.
- Cellular Annotations: Detailed information for each detected cell.

## Cell Annotation Details

Each cell annotation typically includes:

- Bounding Box Coordinates: Defines the spatial extent of the detected cell.
- Centroid Location: Specifies the central point of the cell.
- Classified Cell Type: Assigns a cell type based on the model's prediction.

## GeoJSON File Characteristics

The GeoJSON file complements the JSON data by:

- Providing Exact Geometric Coordinates: Captures precise spatial boundaries of cells.
- Representing Cell Locations as Geospatial Features: Enables integration with geospatial analysis tools.
- Enabling Precise Spatial Analysis and Visualization: Facilitates detailed inspection and visualization of cell locations.

## Visualisation of Cell Types on our web application

Cell detection and classification across Whole Slide Images (WSIs) was achieved by using the Cell-ViT model. The inference workflow comprised patch level analysis with specified imaging parameters, systematic WSI scanning with a tile based methodology, and accurate annotation of various cell types.

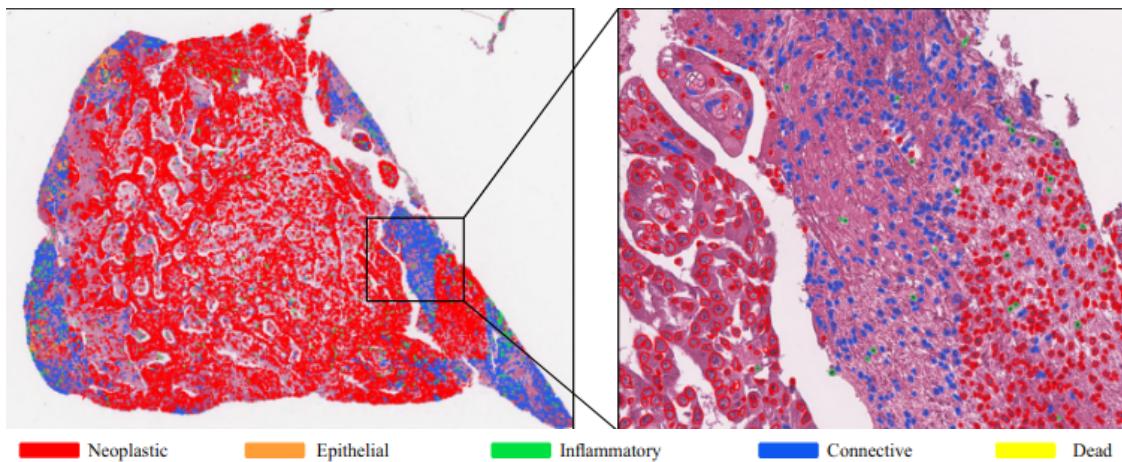


Figure 12: The WSI image shows approximately 150,000 nuclei detected in an exemplary esophageal adenocarcinoma tissue slide at 40x magnification.

## Results from CellViT and CellViT-Plus Model

WSI: 20190610\_100\_401-15\_4350-15-2\_Biopsy\_ER\_HnE\_40X.tif

The file is the TNBC WSI as part of IISER\_batch\_01 dataset. The below figures contains WSI

snapshot the detected cells and matching cells using the CellViT and CellViT-Plus Models. It also includes stats based on their cell types.

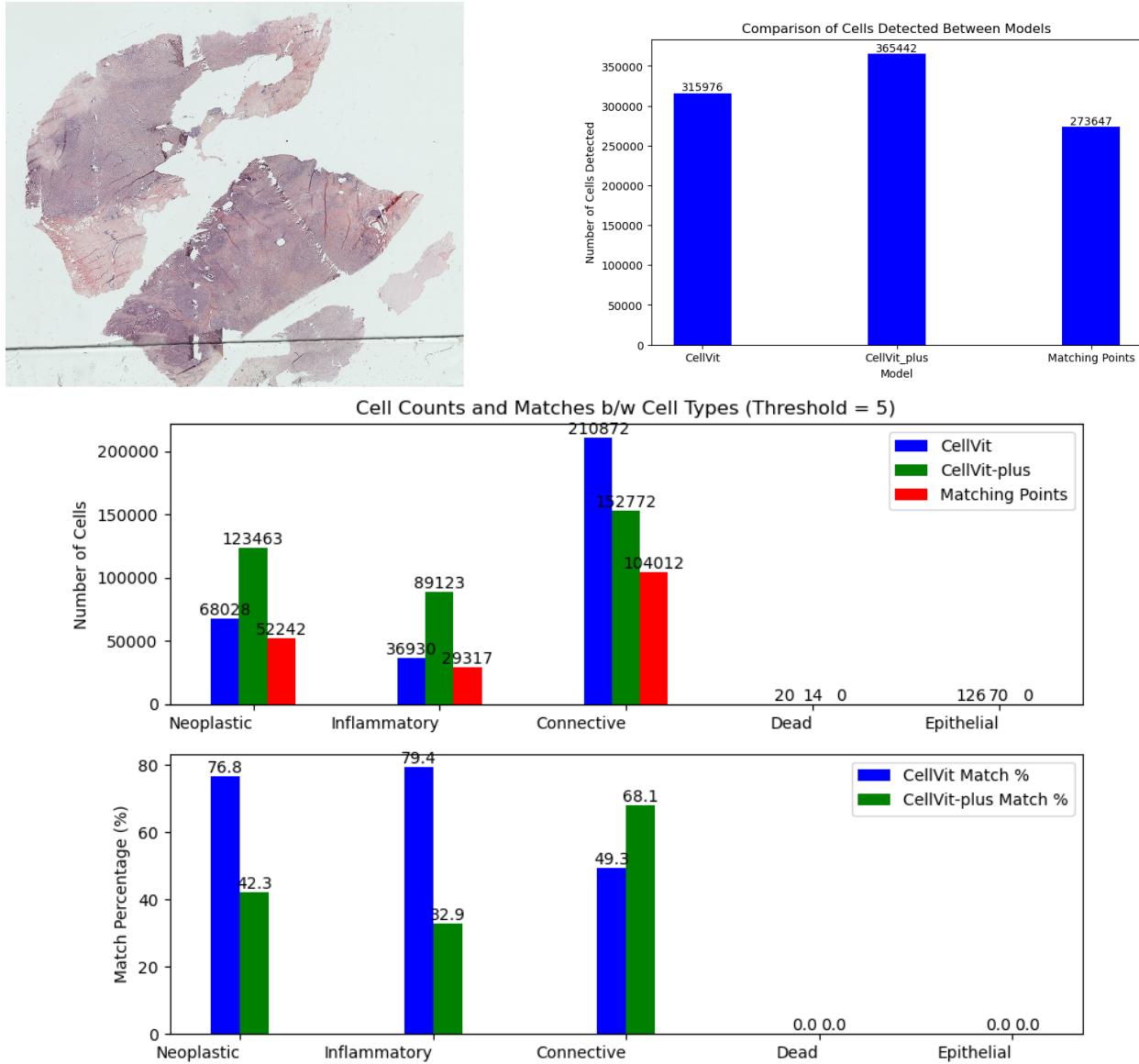


Figure 13: Analysis on IISER batch-01 Whole Slide Image

## 4 Analysis on Gall Bladder Whole Slide Image from AIIMS

A comparative analysis was performed on the gallbladder WSI got from AIIMS. The analysis attached below contains the detected cells and matching cells found using the CellViT and CellViT-Plus Models. It also includes stats based on the cell-types.

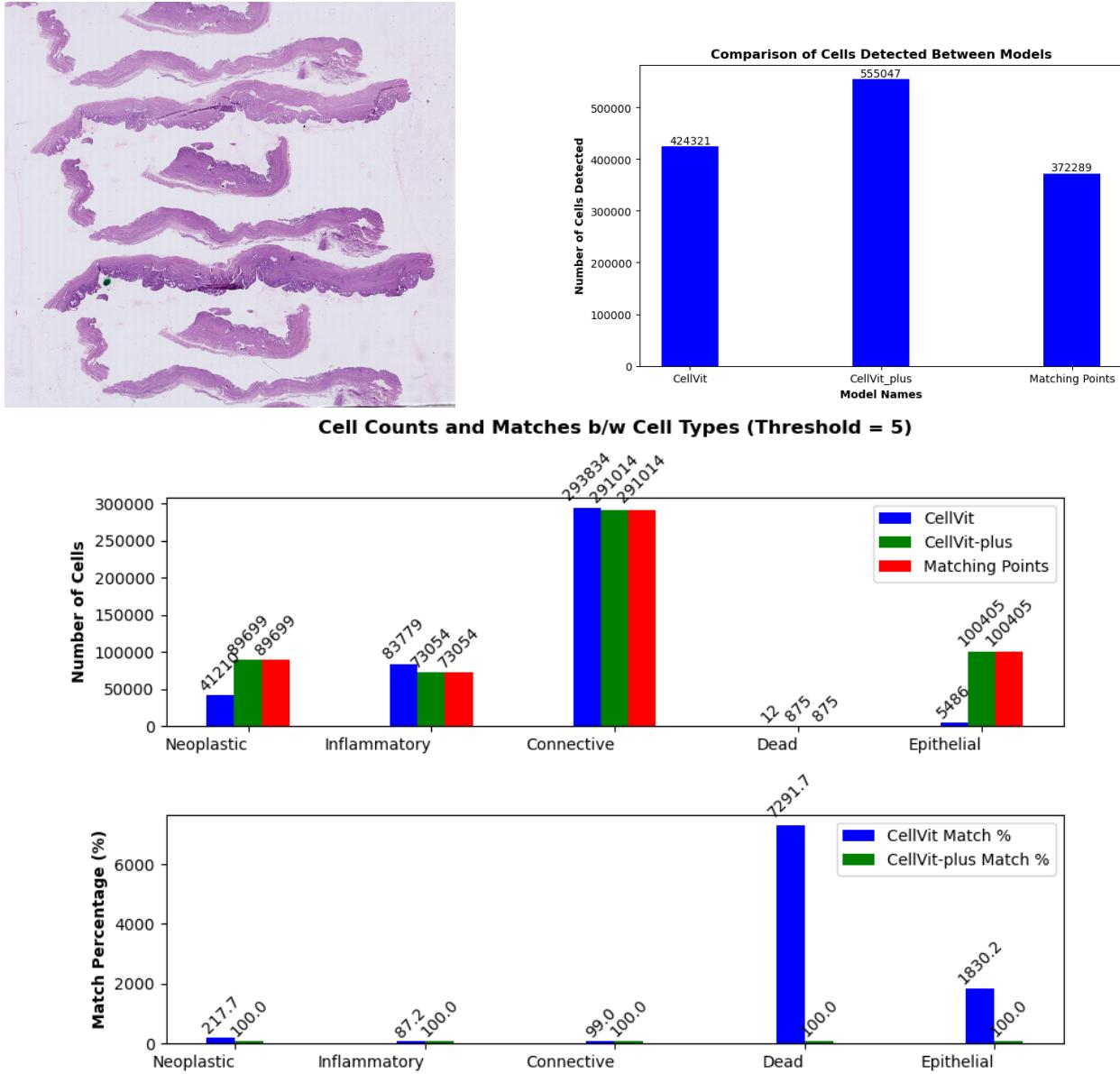


Figure 14: Analysis on AIIMS Gallbladder Whole Slide Image

## Comparative Nuclei Segmentation on AIIMS Gallbladder WSI patch

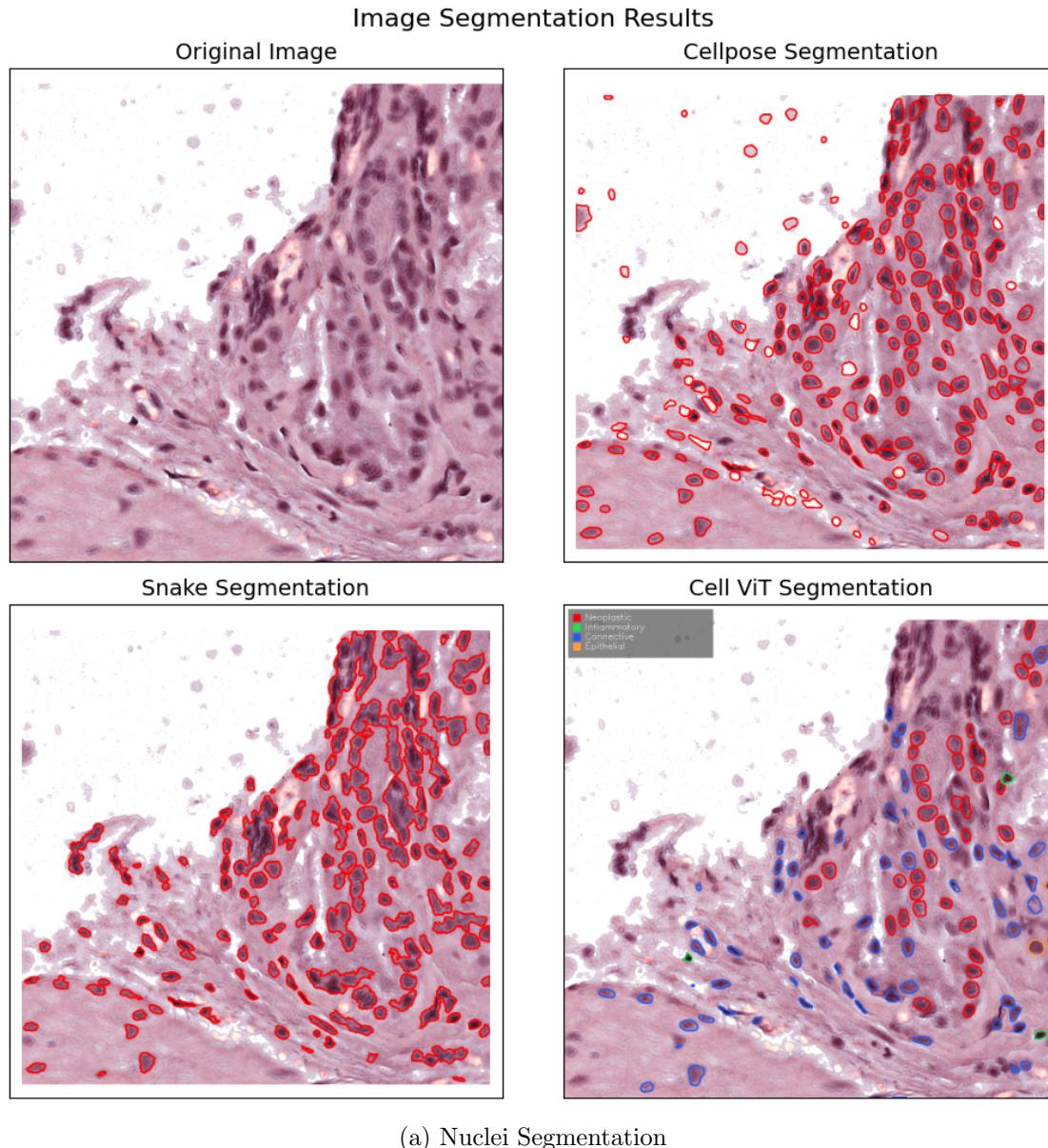
In addition to the work on breast cancer WSIs, an exploratory analysis was conducted to compare the performance of different algorithms for nuclei segmentation on a distinct tissue type. Specifically, a tissue sample from a gallbladder was used to evaluate and visualize the segmentation capabilities of three different computational approaches:

1. **Cellpose:** A deep learning-based segmentation algorithm, generally known for its effectiveness in segmenting various cell types, including nuclei, across diverse microscopy images.
2. **Snake (Active Contour Model):** A classical image segmentation technique where an

initial contour evolves iteratively to fit the boundaries of objects (in this case, nuclei) based on image gradients and other constraints.

3. **CellViT:** The vision transformer-based model previously utilized for cell detection and classification was also applied here to assess its nuclei segmentation performance on this gallbladder tissue sample.

The objective of this comparative analysis was to observe how each of these distinct algorithms delineates nuclei within the same histological context. The segmentation results from Cellpose, the Snake algorithm, and CellViT on the selected gallbladder tissue sample have been compiled and are presented below.



## Comparative Analysis of IHC Markers Across TNBC Whole Slide Images

### 4.1 CellViT and CellViT-Plus Inference results on each marker of H-2992\_18 Tissue Sample

#### IHC Marker 1: CK5

Cytokeratin 5 (CK5) are proteins that provide structure to normal skin and other lining tissues. When found in breast cancer cells (and ER, PR, HER2 are absent), CK5 often indicates a more aggressive type of TNBC called basal-like.

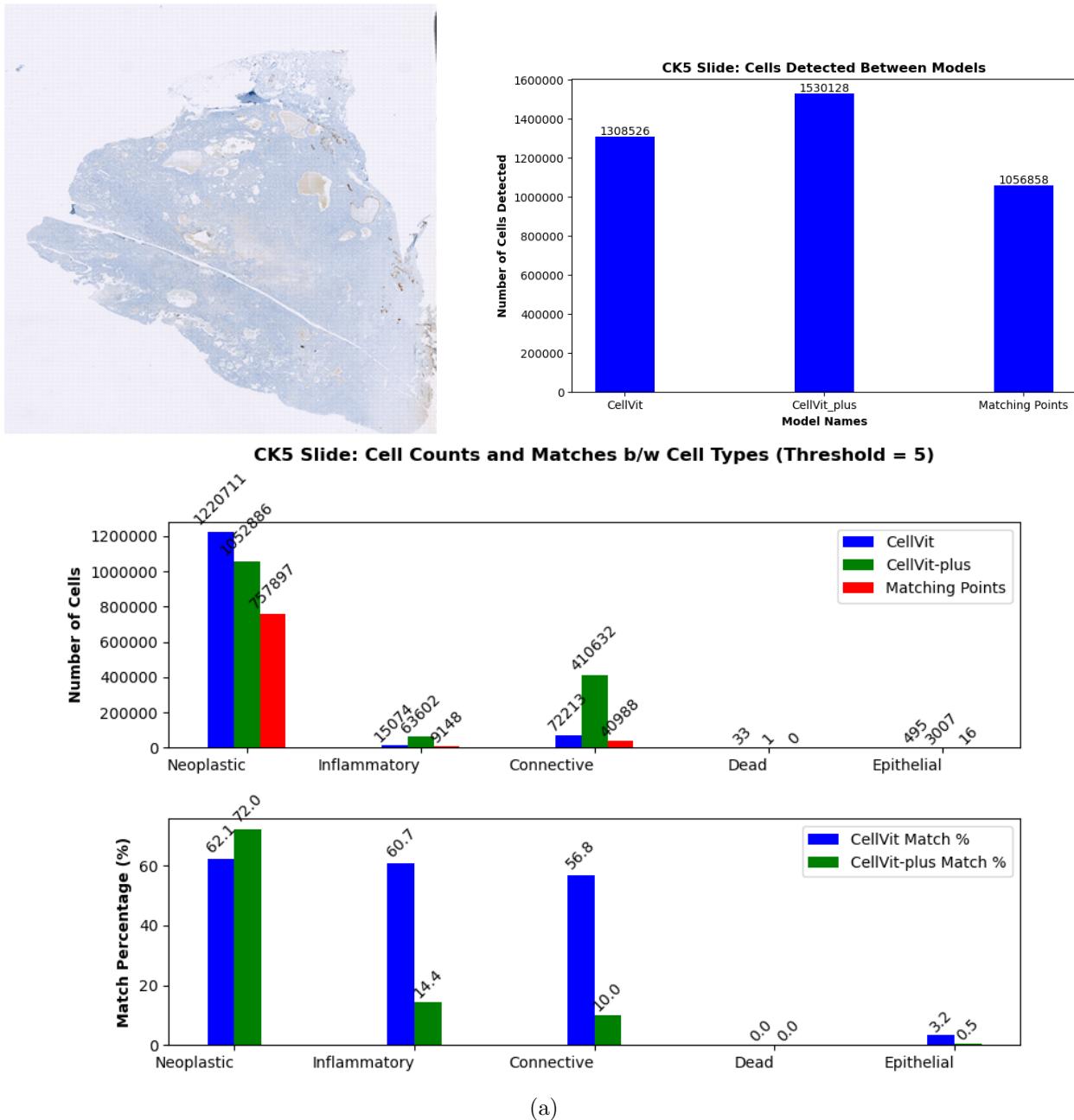
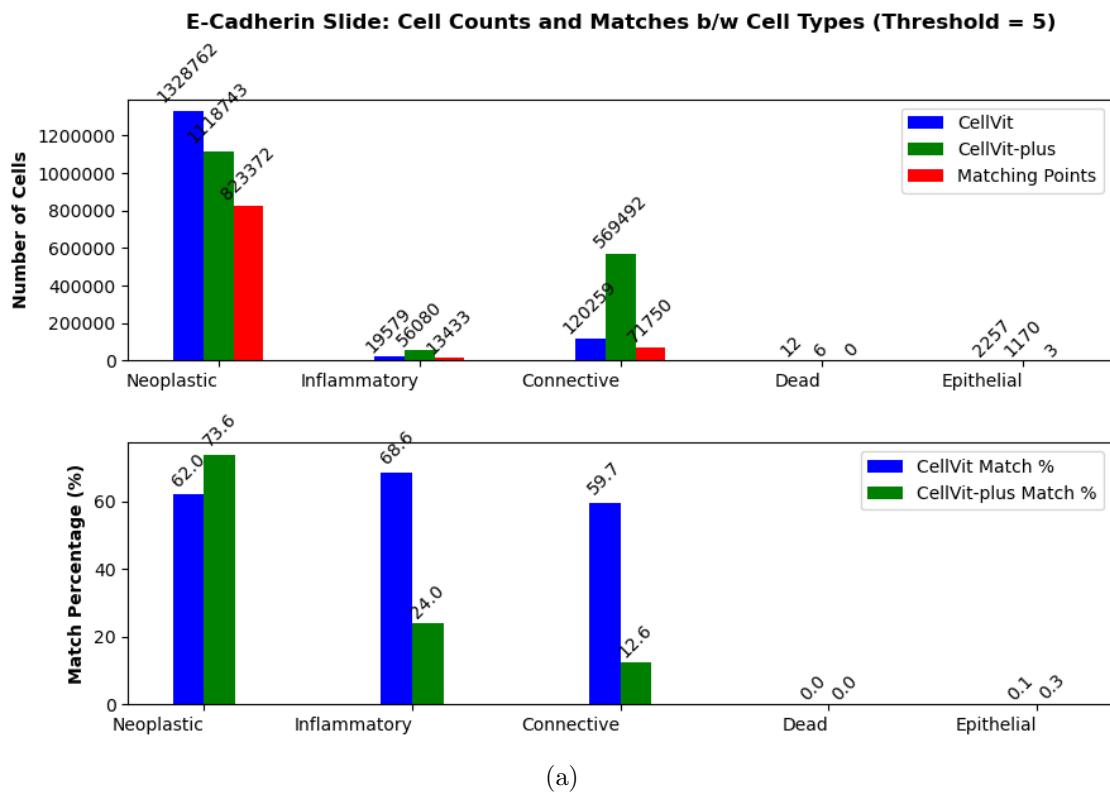
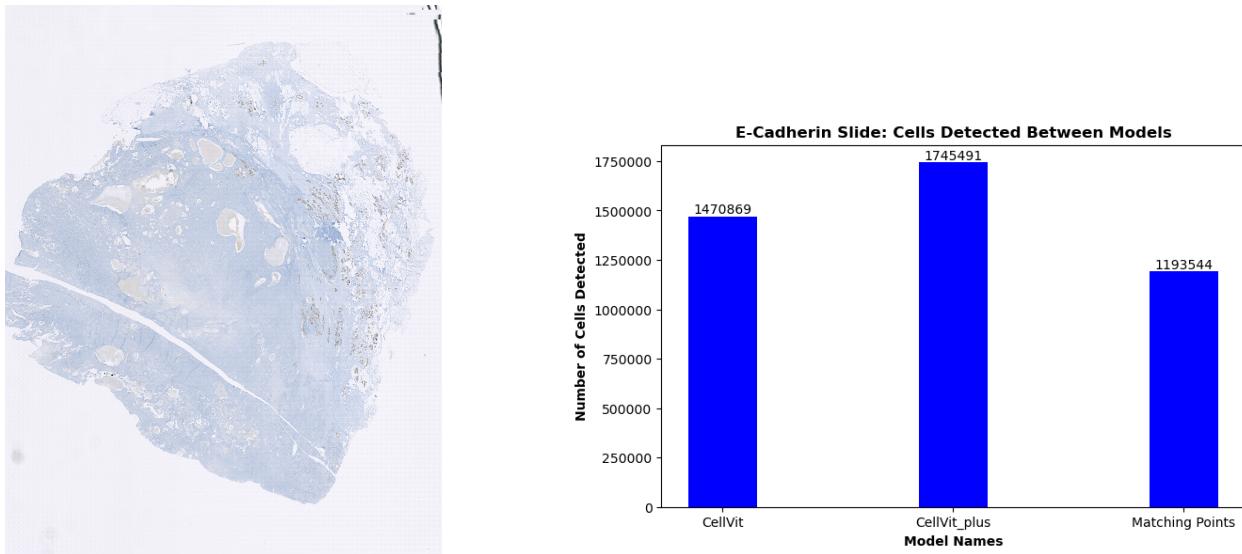


Figure 16: Analysis on CK5 IHC Marker

### IHC Marker 2: E-cadherin

A protein that acts like ‘glue’ holding normal epithelial cells (cells lining surfaces) together. In many cancers, including TNBC, the amount of E-cadherin can decrease or the protein can be lost. Loss of E-cadherin is linked to:

- Increased ability of cancer cells to move and spread
- Poorer survival rates



(a)

Figure 17: Analysis on E-Cadherin IHC Marker

#### IHC Marker 4: Claudin-3

This protein is a key component of tight junctions . Low levels of Claudin-3 are seen in the aggressive claudin low subtype. Interestingly, high levels in the cytoplasm (inside the cell, not on the membrane) have been linked to poorer survival. Claudin-3 is involved in cell-cell adhesion and barrier function and potentially promoting cancer cell movement.

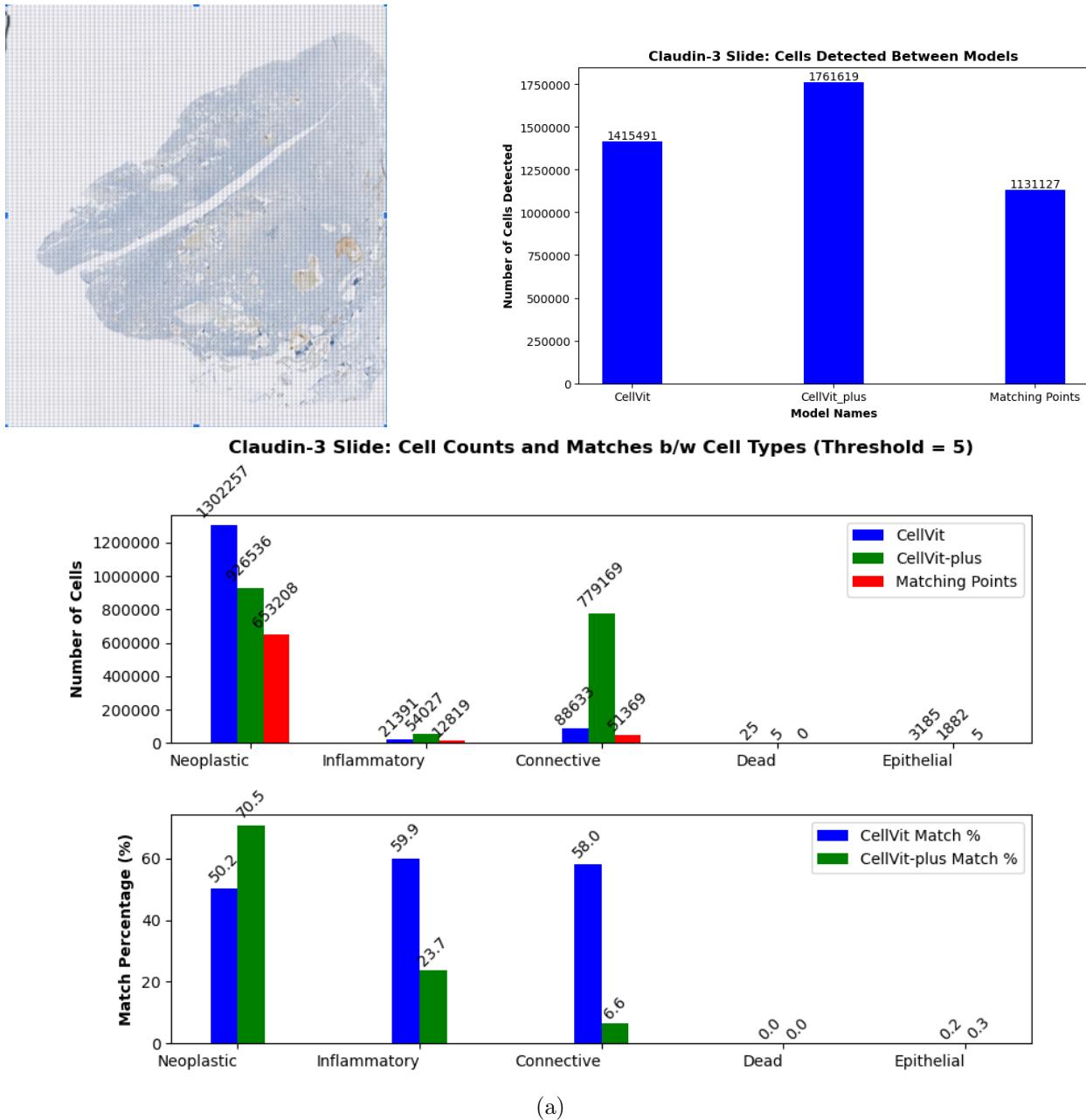


Figure 18: Analysis on Claudin 3 IHC Marker

### IHC Marker 3: Claudin-7

A protein that is a key part of tight junctions, which control what passes between cells in linings like skin and organs. Think of them as selective gates. Claudin-7 levels can be low in an aggressive TNBC subtype called claudin-low. However, its role can be complex, with high levels sometimes linked to more aggressive tumors.

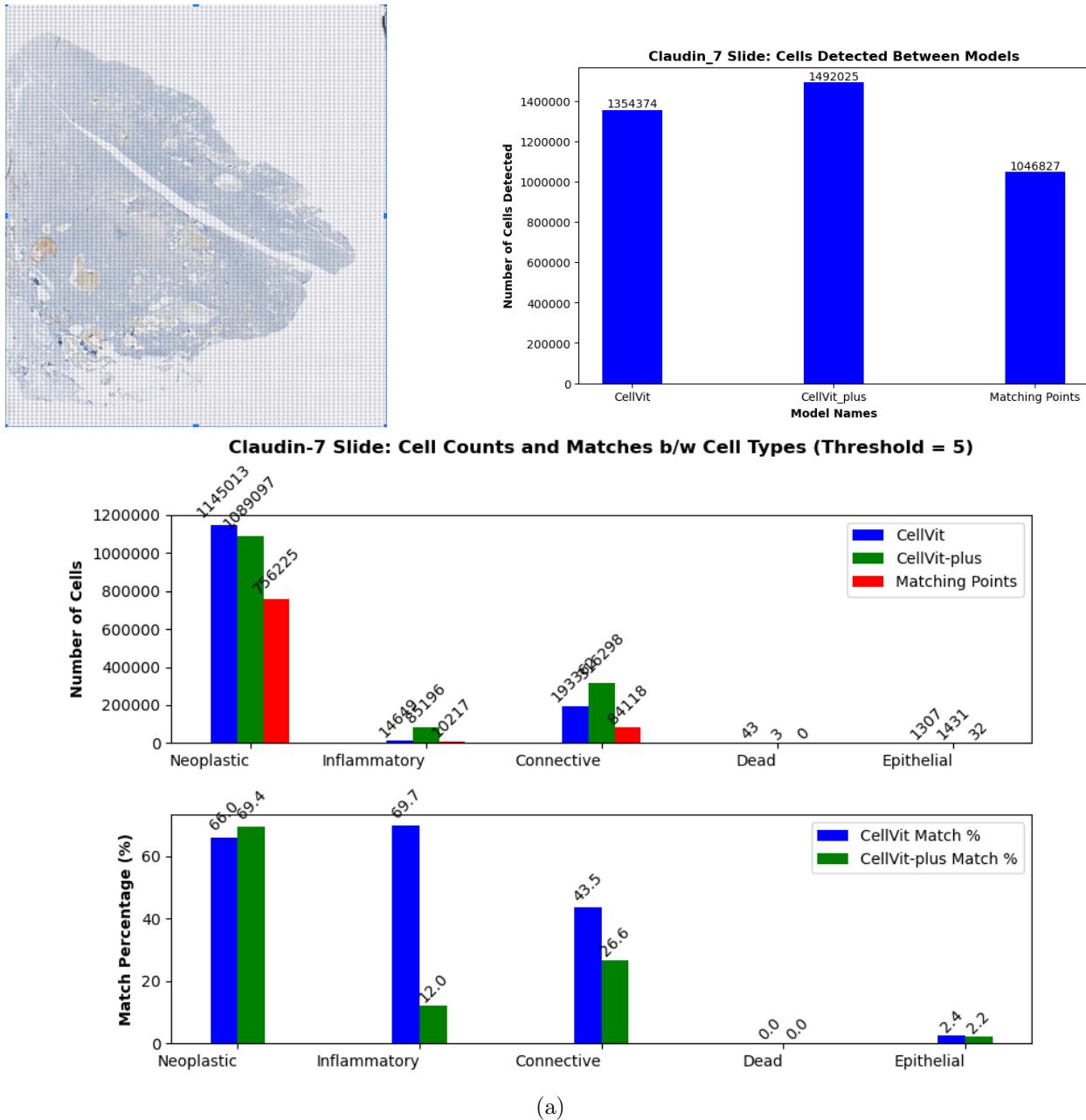


Figure 19: Analysis on Claudin 7 IHC Marker

### IHC Marker 5: Androgen Receptor (AR)

A protein that binds to male hormones (androgens) and controls the activity of certain genes. Some TNBC tumors have AR. This defines a subtype called luminal androgen receptor (LAR), which sometimes has a better prognosis. AR can also be a target for specific therapies.

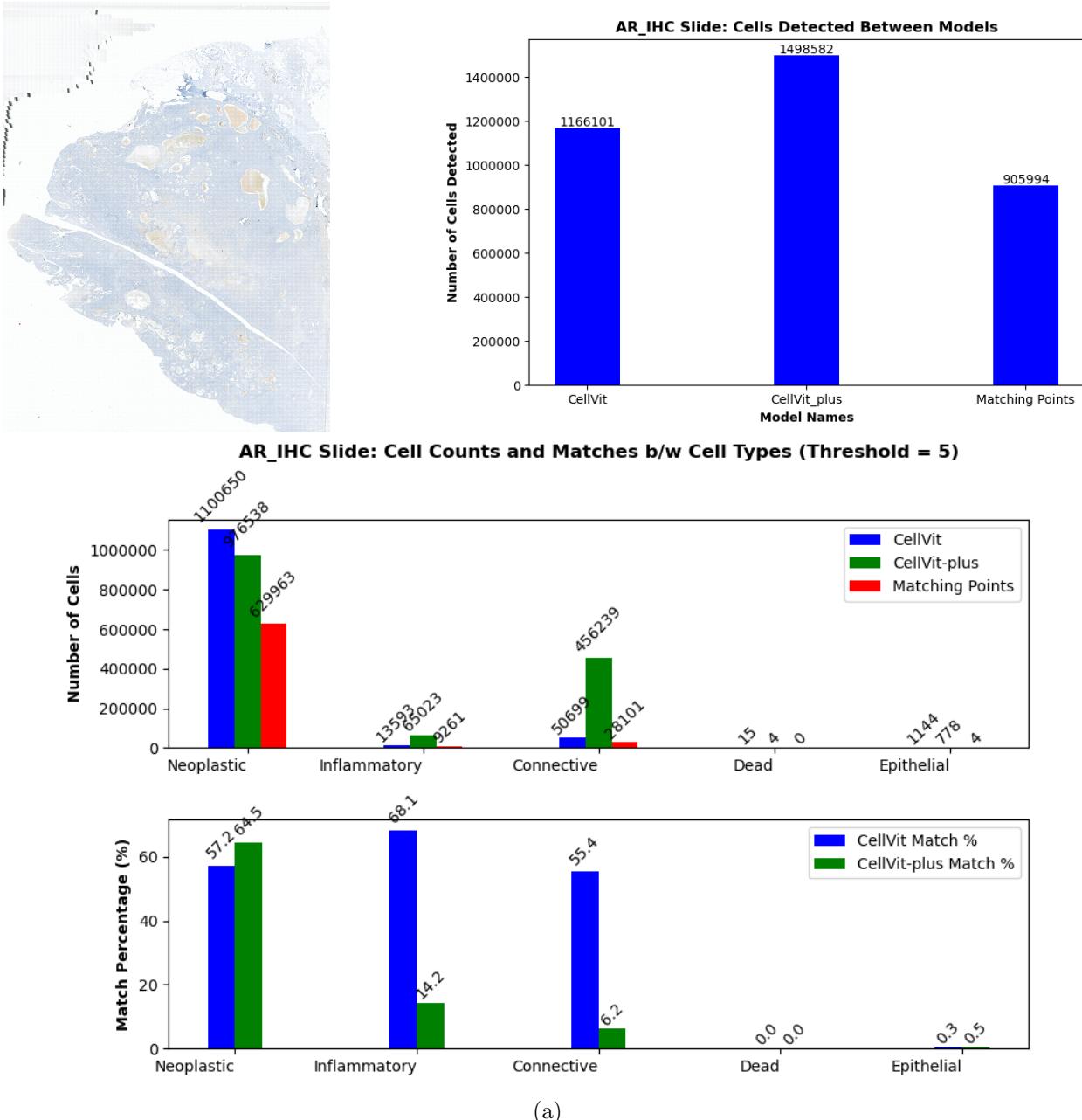
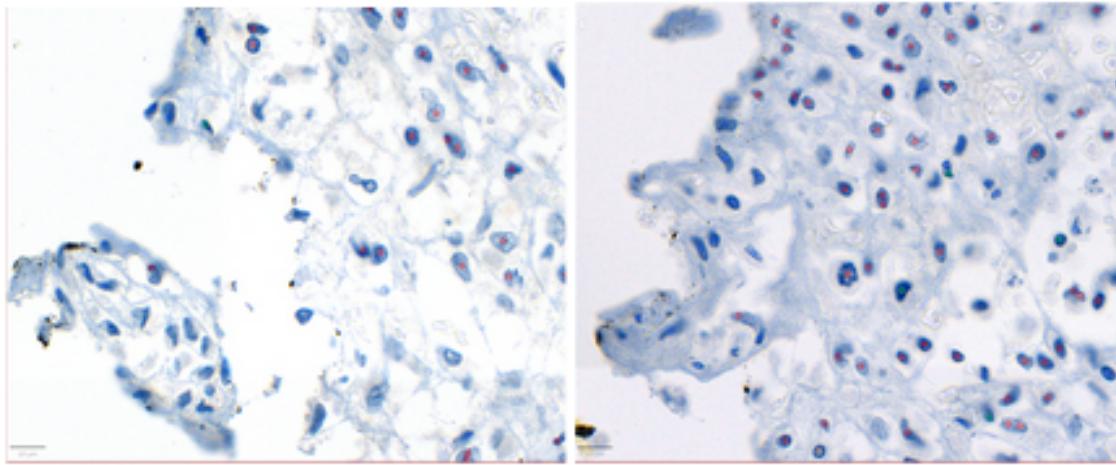


Figure 20: Analysis on AR IHC Marker

### 3) Possible Reason for Difference in Number of Cells Detected in same sample

#### i) Tissue Preparation and Artifact Differences

Although the tissue is from the same sample, the preparation process for each marker has introduced structural changes in the same region of WSI. Differences in cell density and background staining.



(a) Left: AR\_IHC\_slide\_snapshot\_1; Right: E\_Cadherin\_IHC\_slide\_snapshot\_1

## ii) Staining Variability

The AR IHC and E-Cadherin IHC markers target different proteins, leading to distinct staining patterns. This affects how CellViT identifies and classifies cells.

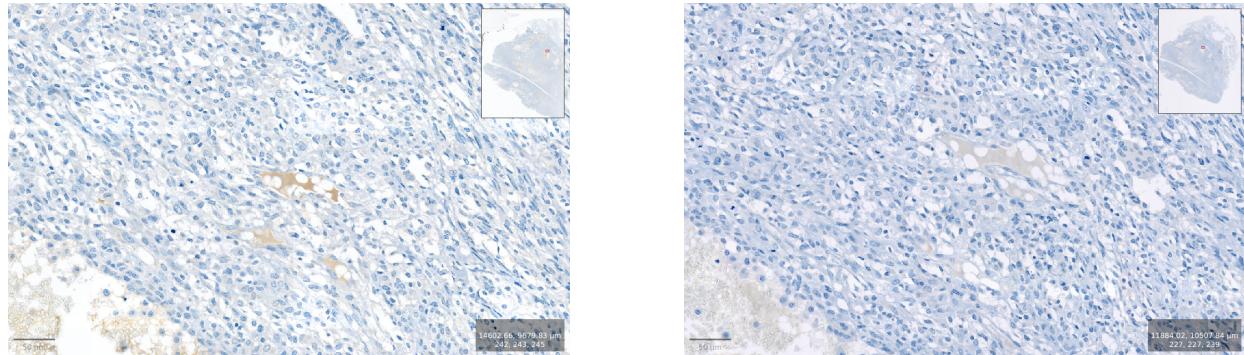
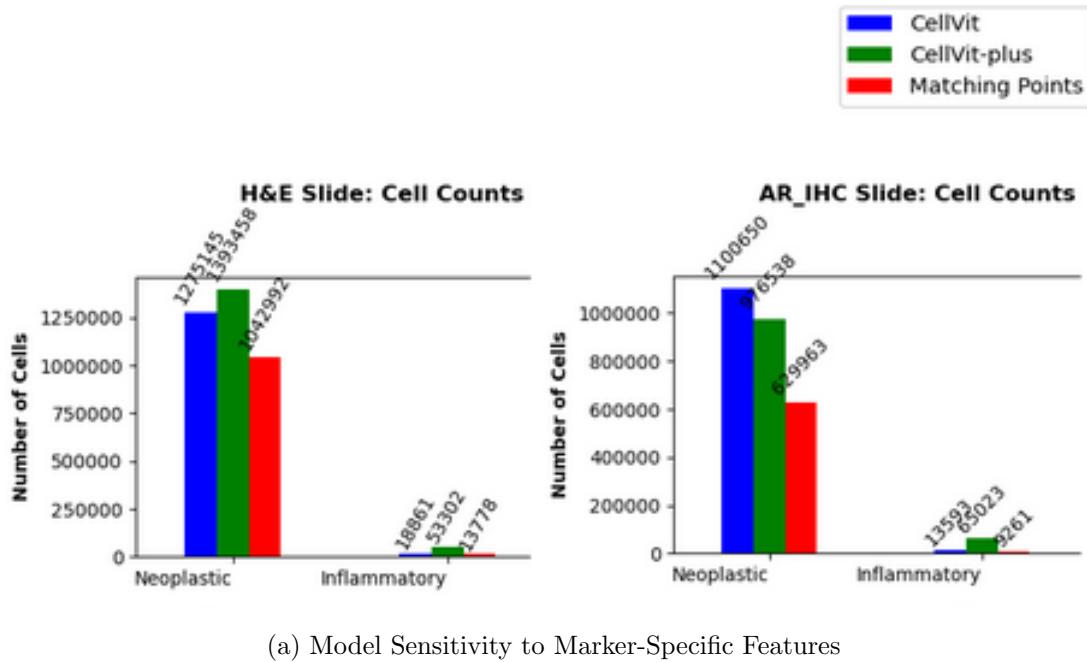


Figure 22: Left: AR\_IHC\_slide\_snapshot\_2; Right: E\_Cadherin\_IHC\_slide\_snapshot\_2

## iii) Model Sensitivity to Marker-Specific Features

The CellViT and CellViT-Plus model is trained upon the PanNuke dataset which contains only Hematoxylin and Eosin (H&E) stained images. For instance, the higher neoplastic cell count in the H&E slide (12,75,145 vs. 11,00,650 for AR) reflects the model's greater sensitivity to H&E epithelial structures and bias towards the training data.



(a) Model Sensitivity to Marker-Specific Features

## 4.2 Data Transfer Process

This part includes the ongoing work that focuses on the secure and efficient transfer of this comprehensive dataset from Ashoka University to Indian Institutes of Science Education and Research (IISER), Pune.

The dataset is currently stored in the cloud. Below are the details.

System Information : RF Grant-2.

IP Address : 10.1.23.173

Dataset path : /media/20TB/augmented-health-system/tnbc/tnbc-max

The dataset includes various immunohistochemical (IHC) markers including :

- E-cadherin
- Claudin-3
- Claudin-7
- Androgen Receptor (AR)
- Cytokeratin 5 (CK-5)
- Hematoxylin and Eosin (H&E)

## 5 Conclusion

This study centered on the evaluation of vision transformer models, with a specific emphasis on Cell-ViT and CellViT-Plus, for the analysis of gigapixel histopathology images, primarily focusing on Triple Negative Breast Cancer (TNBC). The project began with a thorough characterization

of the extensive IISER TNBC dataset. This involved detailed statistical analysis of Whole Slide Images (WSIs) across various experimental rounds, molecular subtypes (ER, HER2, TNBC), and histological/immunohistochemical stains (HnE, Ki-67, CD31, AR, Vimentin, YAP, etc.), alongside essential data harmonization tasks such as the conversion of raw WSI data into a standardized pyramidal TIFF format suitable for high-performance image processing.

The investigation into segmentation techniques progressed from an exploration of foundational methods, including a custom implementation of the Watershed algorithm and an assessment of the Mask R-CNN model, to a comprehensive application of the Cell-ViT architecture. A systematic pipeline was established for WSI processing. This pipeline encompassed automated tissue masking to exclude background regions, standardized patch extraction ( $1024 \times 1024$  pixels with a 64-pixel overlap), stain normalization to address variability, and inference using the Cell-ViT models. This process successfully generated detailed cell-level annotations, including predictions for various cell types such as Neoplastic, Epithelial, Inflammatory, Connective, and Dead cells, along with their corresponding embeddings.

The capabilities of Cell-ViT and CellViT-Plus were demonstrated through their application to H&E stained TNBC WSIs from the IISER dataset. The evaluation was further extended to gallbladder WSIs from AIIMS, where Cell-ViT's segmentation performance was also contextually compared with other algorithms like Cellpose and Active Contour models on selected patches. A significant component of this work involved analyzing Cell-ViT's performance across a panel of diverse Immunohistochemical (IHC) markers (including CK5, E-cadherin, Claudin-3, Claudin-7, and AR) on TNBC tissue samples. This analysis revealed variations in cell detection and classification, influenced by differing staining characteristics and inherent tissue preparation artifacts. These findings highlighted the model's sensitivity, particularly considering its pre-training primarily on H&E stained images from datasets like PanNuke, and underscored the challenges in achieving consistent performance across varied staining protocols without specific fine-tuning or adaptations.

Key outcomes of this ISM include the successful application and rigorous evaluation of the Cell-ViT model in the context of TNBC histopathology, the generation of extensive cell-level annotated datasets, and the derivation of valuable insights into the practical considerations of deploying advanced deep learning models in computational pathology. The analyses provided a clearer understanding of model behavior when faced with different stains and tissue types, emphasizing the critical importance of robust data preprocessing and awareness of potential training data biases. The project efforts also encompassed the preparation and documentation for the secure transfer of the processed and annotated dataset to facilitate ongoing and future collaborative research with IISER, Pune. Overall, this ISM provided a comprehensive evaluation of vision transformer models in histopathology, delivering practical insights, generating valuable annotated datasets, and contributing to the groundwork for further research in AI-driven cancer diagnostics.